Minimizing Markov chains
Beyond Bisimilarity*

Giovanni Bacci, Giorgio Bacci, Kim G. Larsen, Radu Mardare
Aalborg University, Denmark

22 April 2017 - Uppsala, Sweden

SynCoP + PV 2017

(*) On the Metric-based Approximate Minimization of Markov Chains - accepted for ICALP 2017
Best Approximant & Parameter Synthesis

MC(5)
Best Approximant & Parameter Synthesis
Best Approximant & Parameter Synthesis

MC(5)
Best Approximant & Parameter Synthesis
Best Approximant & Parameter Synthesis

MC(5)
Optimal parameters may be irrational
Optimal parameters may be irrational!

\[ x = \frac{1}{30} \left( 10 + \sqrt{163} \right) \]

\[ y = \frac{21}{200} \]
Optimal parameters may be irrational!

\[
\delta(m_0, n_0) = \frac{436}{675} - \frac{163\sqrt{163}}{13500} \approx 0.49
\]

Optimal distance is irrational!

\[
x = \frac{1}{30} \left(10 + \sqrt{163}\right)
\]

\[
y = \frac{21}{200}
\]
The focus of the talk

- Probabilistic Models (Markov chains)
- Automatic verification (e.g., Model Checking)
  - state space explosion (even after model reduction, symbolic tech., partial-order reduction)
- Still too large: one needs to compromise in the accuracy of the model (introduce an error)
- Our proposal: metric-based state space reduction
Probabilistic Bisimulation

[Larsen & Skou’91]
Probabilistic Bisimulation

[Le Larsen & Skou’91]
Probabilistic Bisimulation

[Larsen & Skou’91]
Probabilistic Bisimulation

[Larsen & Skou’91]
Probabilistic Bisimulation

[Probabilistic Bisimulation][Larsen & Skou’91]
Probabilistic Bisimulation

[Optimal lumping
[Kemeny & Snell’60]]

[Larsen & Skou’91]
Probabilistic Bisimulation

[Larsen & Skou’91]

Optimal lumping
[Kemeny & Snell’60]

Efficient technique
[Derisavi et al.’03]
Probabilistic Bisimulation

[Optimal lumping][Kemeny & Snell’60]

[Efficient technique][Derisavi et al.’03]

...but small variations may prevent aggregation
Probabilistic Bisimulation

[łarsen & Skou’91]

...but small variations may prevent aggregation
Probabilistic Bisimulation

$Larsen & Skou’91$

...but small variations may prevent aggregation
Bisimilarity Distance

\[ M = (M, \tau, \ell, m_0) \]

\[ N = (N, \theta, \alpha, n_0) \]
Bisimilarity Distance

\[ \mathcal{M} = (M, \tau, \ell, m_0) \]

\[ \mathcal{N} = (N, \theta, \alpha, n_0) \]
Bisimilarity Distance

\[ \mathcal{M} = (M, \tau, \ell, m_0) \quad \mathcal{N} = (N, \theta, \alpha, n_0) \]
Bisimilarity Distance

\[ \mathcal{M} = (M, \tau, \ell, m_0) \quad \text{and} \quad \mathcal{N} = (N, \theta, \alpha, n_0) \]
Bisimilarity Distance

\[ \mathcal{M} = (M, \tau, \ell, m_0) \]

\[ \mathcal{N} = (N, \theta, \alpha, n_0) \]
Bisimilarity Distance

\[ M = (M, \tau, \ell, m_0) \quad \text{and} \quad N = (N, \theta, \alpha, n_0) \]

Graphical representation with nodes labeled as follows:
- \( m_0 \) to \( m_1 \) with edge weight \( \frac{1}{3} \)
- \( m_0 \) to \( m_2 \) with edge weight \( \frac{2}{3} \)
- \( m_1 \) to \( m_2 \) with edge weight \( \frac{1}{3} \)
- \( m_1 \) to \( m_0 \) with edge weight \( \frac{1}{3} \)
- \( n_0 \) to \( n_1 \) with edge weight \( \frac{1}{3} + \epsilon \)
- \( n_1 \) to \( n_0 \) with edge weight \( \frac{1}{3} \)
- \( n_0 \) to \( n_2 \) with edge weight \( \frac{1}{3} \)
- \( n_1 \) to \( n_2 \) with edge weight \( \epsilon \)
- \( n_2 \) to \( n_3 \) with edge weight \( \frac{1}{3} \)
- \( n_3 \) to \( n_2 \) with edge weight \( \frac{1}{3} - \epsilon \)
- \( n_2 \) to \( n_1 \) with edge weight \( \frac{1}{3} - \epsilon \)
- \( n_1 \) to \( n_2 \) with edge weight \( \epsilon \)
Bisimilarity Distance
(fixed point characterization by van Breugel & Worrell)

Given a parameter $\lambda \in (0, 1]$, called discount factor, the bisimilarity distance $\delta_\lambda$ is the smallest distance satisfying

$$
\delta_\lambda(m,n) = \begin{cases} 
1 & \text{if } \ell(m) \neq \alpha(n) \\
\lambda \cdot \mathcal{K}(\delta_\lambda)(\tau(m),\theta(n)) & \text{otherwise}
\end{cases}
$$
Bisimilarity Distance

(fixed point characterization by van Breugel & Worrell)

Given a parameter \( \lambda \in (0, 1] \), called \textit{discount factor}, the \textit{bisimilarity distance} \( \delta_\lambda \) is the smallest distance satisfying

\[
\delta_\lambda(m,n) = \begin{cases} 
1 & \text{if } \ell(m) \neq \alpha(n) \\
\lambda \cdot \mathcal{K}(\delta_\lambda)(\tau(m),\theta(n)) & \text{otherwise}
\end{cases}
\]

\textbf{Kantorovich lifting}
Bisimilarity Distance

(fixed point characterization by van Breugel & Worrell)

Given a parameter \( \lambda \in (0, 1] \), called discount factor, the **bisimilarity distance** \( \delta_\lambda \) is the smallest distance satisfying

\[
\delta_\lambda(m, n) = \begin{cases} 
1 & \text{if } \ell(m) \neq \alpha(n) \\
\lambda \cdot \mathcal{K}(\delta_\lambda)(\tau(m), \theta(n)) & \text{otherwise}
\end{cases}
\]

\[
\mathcal{K}(d)(\tau(m), \theta(n)) = \min \left\{ \sum d(u, v) \cdot C(u, v) \left| \begin{array}{l}
\sum_{u \in M} C(u, v) = \theta(n)(v) \\
\sum_{v \in N} C(u, v) = \tau(m)(u)
\end{array} \right. \right\}
\]
Bisimilarity Distance
(fixed point characterization by van Breugel & Worrell)

Given a parameter $\lambda \in (0,1]$, called discount factor, the bisimilarity distance $\delta_\lambda$ is the smallest distance satisfying

$$\delta_\lambda(m,n) = \begin{cases} 1 & \text{if } \ell(m) \neq \alpha(n) \\ \lambda \cdot \mathcal{K}(\delta_\lambda)(\tau(m),\theta(n)) & \text{otherwise} \end{cases}$$

$$\mathcal{K}(d)(\tau(m),\theta(n)) = \min \left\{ \sum d(u,v) \cdot C(u,v) \right\} \quad \left\{ \begin{array}{l} \sum_{u \in M} C(u,v) = \theta(n)(v) \\ \sum_{v \in N} C(u,v) = \tau(m)(u) \end{array} \right\}$$

**Discount at each step**

**Kantorovich lifting**

**Coupling**
Remarkable properties

Theorem (Desharnais et al. 99)

\[ m \sim n \quad \text{iff} \quad \delta_\lambda(m,n) = 0 \]

Theorem (Chen, van Breugel, Worrell 12)

The probabilistic bisimilarity distance can be computed in polynomial time

(Jonsson & L 91)

(Bacci, L, Mardare 13)
Approximate verification

Theorem (Chen et al. FoSSaCS’12, Bacci et al. ICTAC’15)

\[|P(\mathcal{M})([\varphi]) - P(\mathcal{N})([\varphi])| \leq \delta_1(\mathcal{M}, \mathcal{N})\]

for all LTL formulas!

difference in the probability of satisfying \(\varphi\)
Approximate verification

**Theorem (Chen et al. FoSSaCS’12, Bacci et al. ICTAC’15)**

\[
|P(\mathcal{M})([\varphi]) - P(\mathcal{N})([\varphi])| \leq \delta_1(\mathcal{M},\mathcal{N})
\]

for all LTL formulas!

…imagine that \(|\mathcal{M}| \gg |\mathcal{N}|\), we can use \(\mathcal{N}\) in place of \(\mathcal{M}\)

\[P(\mathcal{N})([\varphi])\quad \text{approximate solution on } \varphi\]

\[P(\mathcal{M})([\varphi])\]
Metric-based State Space Reduction

Closest Bounded Approximant (CBA)  Minimum Significant Approximant Bound (MSAB)
Metric-based State Space Reduction

Closest Bounded Approximant (CBA)

Minimum Significant Approximant Bound (MSAB)

minimize $d$
Metric-based State Space Reduction

Closest Bounded Approximant (CBA)

$$\text{minimize } d \quad \text{MC}(k)$$

Minimum Significant Approximant Bound (MSAB)

$$\text{minimize } k$$
List of our Results

- CBA as bilinear program
- The CBA’s threshold problem is NP-hard (complexity lower bound)
  - PSPACE (complexity upper bound)
- The MSAB’s threshold problem is NP-complete
- Expectation Maximization heuristic for CBA
The CBA-$\lambda$ problem

The Closest Bounded Approximant w.r.t. $\delta_\lambda$

**Instance:** An MC $M$, and a positive integer $k$

**Output:** An MC $\tilde{N}$, with at most $k$ states
minimizing $\delta_\lambda(m_0,\tilde{n}_0)$

$$\delta_\lambda(m_0,\tilde{n}_0) = \inf \{ \delta_\lambda(m_0,n_0) \mid N \in \text{MC}(k) \}$$

we get a solution iff the infimum is a minimum
The CBA-\(\lambda\) problem

The Closest Bounded Approximant w.r.t. \(\delta_\lambda\)

\textbf{Instance:} An MC \(M\), and a positive integer \(k\)

\textbf{Output:} An MC \(\tilde{N}\), with at most \(k\) states

\(\delta_\lambda(m_0,\tilde{n}_0) = \inf \{ \delta_\lambda(m_0,n_0) \mid N \in \text{MC}(k) \} \)

we get a solution iff the infimum is a minimum

generalization of bisimilarity quotient
CBA-λ as a Bilinear Program

**Lemma (Meaningful labels)**

For any $N \in MC(k)$, there exists $N' \in MC(k)$ with labels taken from $M$, such that $\delta_{\lambda}(M,N) \geq \delta_{\lambda}(M,N')$
CBA-\(\lambda\) as a Bilinear Program

**Lemma (Meaningful labels)**

For any \(N \in \text{MC}(k)\), there exists \(N' \in \text{MC}(k)\) with labels taken from \(M\), such that \(\delta_\lambda(M,N) \geq \delta_\lambda(M,N')\)

minimize \(d_{m_0,n_0}\)

such that \(\lambda \sum_{(u,v) \in M \times N} c_{u,v}^m \cdot d_{u,v} \leq d_{m,n}\)

\(1 - \alpha_{n,l} \leq d_{m,n} \leq 1\)

\(\alpha_{n,l} \cdot \alpha_{n,l'} = 0\)

\(\sum_{l \in L(M)} \alpha_{n,l} = 1\)

\(\sum_{v \in N} c_{u,v}^m = \tau(m)(u)\)

\(\sum_{u \in M} c_{u,v}^m = \theta_{n,v}\)

\(c_{u,v}^m \geq 0\)

\(m \in M, n \in N\)

\(n \in N, l \in L(M), l(m) \neq l\)

\(n \in N, l, l' \in L(M), l \neq l'\)

\(n \in N\)

\(m, u \in M, n \in N\)

\(m \in M, n, v \in N\)

\(m, u \in M, n, v \in N\)
CBA-\(\lambda\) as a Bilinear Program

**Lemma (Meaningful labels)**

For any \(N\in MC(k)\), there exists \(N'\in MC(k)\) with labels taken from \(M\), such that \(\delta_\lambda(M,N) \geq \delta_\lambda(M,N')\)

<table>
<thead>
<tr>
<th>minimize (d_{m_0,n_0})</th>
<th>such that</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda \sum_{(u,v)\in M\times N} c_{u,v}^{m,n} \cdot d_{u,v} \leq d_{m,n})</td>
<td>(m \in M, n \in N)</td>
</tr>
<tr>
<td>(1 - \alpha_{n,l} \leq d_{m,n} \leq 1)</td>
<td>(n \in N, l \in L(M), \ell(m) \neq l)</td>
</tr>
<tr>
<td>(\alpha_{n,l} \cdot \alpha_{n,l'} = 0)</td>
<td>(n \in N, l, l' \in L(M), l \neq l')</td>
</tr>
<tr>
<td>(\sum_{l\in L(M)} \alpha_{n,l} = 1)</td>
<td>(n \in N)</td>
</tr>
<tr>
<td>(\sum_{v\in N} c_{u,v}^{m,n} = \tau(m)(u))</td>
<td>(m, u \in M, n \in N)</td>
</tr>
<tr>
<td>(\sum_{u\in M} c_{u,v}^{m,n} = \theta_{n,v})</td>
<td>(m \in M, n, v \in N)</td>
</tr>
<tr>
<td>(c_{u,v}^{m,n} \geq 0)</td>
<td>(m, u \in M, n, v \in N)</td>
</tr>
</tbody>
</table>
CBA-$\lambda$ as a Bilinear Program

**Lemma (Meaningful labels)**

For any $N \in MC(k)$, there exists $N' \in MC(k)$ with labels taken from $M$, such that $\delta_\lambda(M, N) \geq \delta_\lambda(M, N')$

minimize $d_{m_0, n_0}$
such that $\lambda \sum_{(u, v) \in M \times N} c_{u, v}^{m, n} \cdot d_{u, v} \leq d_{m, n}$
$1 - \alpha_{n, l} \leq d_{m, n} \leq 1$
$\alpha_{n, l} \cdot \alpha_{n, l'} = 0$
$\sum_{l \in L(M)} \alpha_{n, l} = 1$
$\sum_{v \in N} c_{u, v}^{m, n} = \tau(m)(u)$
$\sum_{u \in M} c_{u, v}^{m, n} = \theta_{n, v}$
$c_{u, v}^{m, n} \geq 0$

$m \in M, n \in N$
$n \in N, l \in L(M), \ell(m) \neq l$
$n \in N, l, l' \in L(M), l \neq l'$
$n \in N$
$m, u \in M, n \in N$
$m \in M, n, v \in N$
$m, u \in M, n, v \in N$
CBA-\(\lambda\) as a Bilinear Program

**Lemma (Meaningful labels)**
For any \(N \in MC(k)\), there exists \(N' \in MC(k)\) with labels taken from \(M\), such that \(\delta_\lambda(M,N) \geq \delta_\lambda(M,N')\)

minimize \(d_{m_0,n_0}\) such that \(\lambda \sum_{(u,v) \in M \times N} c_{u,v}^{m,n} \cdot d_{u,v} \leq d_{m,n}\)

\[1 - \alpha_{n,l} \leq d_{m,n} \leq 1\]

\[\alpha_{n,l} \cdot \alpha_{n,l'} = 0\]

\[\sum_{l \in L(M)} \alpha_{n,l} = 1\]

\[\sum_{v \in N} c_{u,v}^{m,n} = \tau(m)(u)\]

\[\sum_{u \in M} c_{u,v}^{m,n} = \theta_{n,v}\]

\[c_{u,v}^{m,n} \geq 0\]
CBA-\(\lambda\) as a Bilinear Program

this characterization has two main consequences…

1. CBA-\(\lambda\) admits always a solution

2. CBA-\(\lambda\) is can be approximated up to any precision
Complexity of CBA-λ

“To study the complexity of an optimization problem one has to look at its decision variant”

(C. Papadimitriou)
Complexity of CBA-$\lambda$

“To study the complexity of an optimization problem one has to look at its decision variant”

(C. Papadimitriou)

**Bounded Approximant threshold wrt $d_\lambda$**

**Instance:** An MC $M$, a positive integer $k$, and a *rational* $\varepsilon > 0$

**Output:** *yes* iff there exists $N$ with at most $k$ states such that $\delta_\lambda(m_0,n_0) \leq \varepsilon$
Complexity upper bound

**Theorem**

BA-\( \lambda \) is in **PSPACE**

*Proof sketch:* we can encode the question \( \langle M, k, \varepsilon \rangle \in \text{BA-} \lambda \) to that of checking the feasibility of a set of bilinear inequalities. This can be encoded as a decision problem for the existential theory of the reals, thus it can be solved in PSPACE [Canny—STOC88].
Complexity lower bound

Theorem

$\mathbf{BA-\lambda}$ is $\mathbf{NP}$-hard

Proof idea: we provide a reduction from VERTEX COVER.
Complexity lower bound

**Theorem**

$\text{BA-}\lambda$ is **NP-hard**

unlikely to solve CBA as simple linear program

**Proof idea:** we provide a reduction from VERTEX COVER.
The MSAB-$\lambda$ problem

The Minimum Significant Approximant Bound wrt $\delta_\lambda$

**Instance:** An MC $M$

**Output:** The smallest $k$ such that $d_\lambda(m_0,n_0)<1$, for some $N\in MC(k)$
The MSAB-$\lambda$ problem

The Minimum Significant Approximant Bound wrt $\delta_{\lambda}$

**Instance:** An MC $M$

**Output:** The smallest $k$ such that $d_{\lambda}(m_0,n_0) < 1$, for some $N \in MC(k)$

For $\lambda < 1$, the MSAB-$\lambda$ problem is trivial, because the solution is always $k = 1$
The MSAB-\(\lambda\) problem

**The Minimum Significant Approximant Bound wrt** \(\delta_{\lambda}\)**

**Instance:** An MC \(M\)

**Output:** The smallest \(k\) such that \(d_{\lambda}(m_0,n_0)<1\), for some \(N\in MC(k)\)

For \(\lambda<1\), the MSAB-\(\lambda\) problem is trivial, because the solution is always \(k=1\)

For \(\lambda=1\), the same problem is surprisingly difficult…
Complexity of MSAB-1

...as before we should look at its decision variant
Complexity of MSAB-1

...as before we should look at its decision variant

**Significant Bounded Approximant wrt $\delta_1$**

**Instance:** An MC $M$ and a **positive** $k$

**Output:** **yes** iff there exists $N$ with **at most** $k$
states such that $\delta_1(m_0,n_0) < 1$. 
Complexity of MSAB-1

...as before we should look at its decision variant

**Significant Bounded Approximant wrt $\delta_1$**

**Instance:** An MC $M$ and a positive $k$

**Output:** yes iff there exists $N$ with at most $k$ states such that $\delta_1(m_0,n_0)<1$.

**Theorem**

SBA-1 is **NP-complete**
Lemma

Assume $M$ be maximally collapsed. Then,

$$\langle M,k \rangle \in SBA-1 \quad \text{iff} \quad G(M) = BSCC \quad \text{and} \quad h + |C| \leq k$$

number of labels in $m_0 \ldots m_{n-1}$
Lemma

Assume M be maximally collapsed. Then,

\[ \langle M, k \rangle \in \text{SBA-1} \quad \text{iff} \quad G(M) = C \quad \text{and} \quad h + |C| \leq k \]

Proof sketch: compute with Tarjan’s algorithm all the SCCs of \( G(M) \). Then non deterministically choose a BSCC and a path to it. In poly-time we can count the number of labels in the path and the size of the BSCC.
SBA-1 is NP-hard

Proof sketch: by reduction to VERTEX COVER:

\[ \langle G, h \rangle \in \text{VERTEX COVER} \iff \langle M_G, h+m+1 \rangle \in \text{SBA-1} \]
SBA-1 is **NP-hard**

**Proof sketch:** by reduction to VERTEX COVER:

\[
\langle G, h \rangle \in \text{VERTEX COVER} \quad \text{iff} \quad \langle M_G, h+m+1 \rangle \in \text{SBA-1}
\]

paths from \(e_3\) to \(e_0\) describes all vertex covers of \(G\)
Towards an Algorithm…
Towards an Algorithm…

• CBA can be solved as a bilinear program. Theoretically nice, but practically unfeasible! (our implementation in PENBMI can handle MCs with at most 5 states…)}
Towards an Algorithm…

• CBA can be solved as a bilinear program. Theoretically nice, but practically unfeasible!
  (our implementation in PENBMI can handle MCs with at most 5 states…)

• We are happy with sub-optimal solutions if they can be obtained by a practical algorithm.
EM-like Algorithm

- Given the MC M and an initial approximant N_0
- it produces a sequence N_0, …, N_h of approximants having strictly decreasing distance from M
- N_h may be a sub-optimal solution of CBA-λ
EM-like Algorithm

Algorithm 1

Input: $\mathcal{M} = (M, \tau, \ell)$, $\mathcal{N}_0 = (N, \theta_0, \alpha)$, and $h \in \mathbb{N}$.

1. $i \leftarrow 0$
2. repeat
3. $i \leftarrow i + 1$
4. compute $C \in \Omega(\mathcal{M}, \mathcal{N}_{i-1})$ such that $\delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1}) = \gamma_\lambda^C(\mathcal{M}, \mathcal{N}_{i-1})$
5. $\theta_i \leftarrow \text{UPDATETRANSITION}(\theta_{i-1}, C)$
6. $\mathcal{N}_i \leftarrow (N, \theta_i, \alpha)$
7. until $\delta_\lambda(\mathcal{M}, \mathcal{N}_i) > \delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1})$ or $i \geq h$
8. return $\mathcal{N}_{i-1}$
EM-like Algorithm

Algorithm 1

Input: $\mathcal{M} = (M, \tau, \ell)$, $\mathcal{N}_0 = (N, \theta_0, \alpha)$, and $h \in \mathbb{N}$.

1. $i \leftarrow 0$
2. repeat
3. $i \leftarrow i + 1$
4. compute $\mathcal{C} \in \Omega(\mathcal{M}, \mathcal{N}_{i-1})$ such that $\delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1}) = \gamma_\lambda^\mathcal{C}(\mathcal{M}, \mathcal{N}_{i-1})$
5. $\theta_i \leftarrow \text{UPDATETRANSITION}(\theta_{i-1}, \mathcal{C})$
6. $\mathcal{N}_i \leftarrow (N, \theta_i, \alpha)$
7. until $\delta_\lambda(\mathcal{M}, \mathcal{N}_i) > \delta_\lambda(\mathcal{M}, \mathcal{N}_{i-1})$ or $i \geq h$
8. return $\mathcal{N}_{i-1}$

Intuitive Idea

UpdateTransition assigns greater probability to transitions that are most representative of the behavior of $M$. 
Two update heuristics

- **Averaged Marginal (AM):** given $N_k$ we construct $N_{k+1}$ by averaging the marginal of certain “coupling variables” obtained by optimizing the number of occurrences of the edges that are most likely to be seen in $M$.

- **Averaged Expectations (AE):** similar to the above, but now the $N_{k+1}$ looks only the expectation of the number of occurrences of the edges likely to be found in $M$. 
IPv4 Zero Conf Protocol

Averaged Marginal (AM)
IPv4 Zero Conf Protocol

Averaged Marginal (AM)

$\delta_{0.9}(M,N_0) \approx 0.67$
IPv4 Zero Conf Protocol

Averaged Marginal (AM)

\[ \delta_{0.9}(M,N_0) \approx 0.67 \]

\[ \delta_{0.9}(M,N_1) \approx 0.043 \]
IPv4 Zero Conf Protocol

Averaged Marginal (AM)

\[ \delta_{0.9}(M,N_0) \approx 0.67 \]

\[ \delta_{0.9}(M,N_1) \approx 0.043 \]

\[ \delta_{0.9}(M,N_2) \approx 0.041 \]
IPv4 Zero Conf Protocol

Averaged Expectations (AE)

$$\delta_{0.9}(M,N_0) \approx 0.67$$
IPv4 Zero Conf Protocol

Averaged Expectations (AE)

$$\delta_{0.9}(M,N_0) \approx 0.67$$

$$\delta_{0.9}(M,N_1) \approx 0.08$$
IPv4 Zero Conf Protocol

Averaged Expectations (AE)

\[ \delta_{0.9}(M, N_0) \approx 0.67 \]

\[ \delta_{0.9}(M, N_1) \approx 0.08 \]

\[ \delta_{0.9}(M, N_2) \approx 0.11 \]
Drunkard's Walk

Averaged Marginal (AM)
Drunkard's Walk

Averaged Marginal (AM)

\[ \delta_{0.9}(M,N_0) \approx 0.64 \]
Drunkard's Walk

Averaged Marginal (AM)

\[
\delta_{0.9}(M, N_0) \approx 0.64
\]

\[
\delta_{0.9}(M, N_1) \approx 0.56
\]
Drunkard's Walk

Averaged Marginal (AM)

\[ \delta_{0.9}(M, N_0) \approx 0.64 \]

\[ \delta_{0.9}(M, N_1) \approx 0.56 \]

\[ \delta_{0.9}(M, N_2) \approx 0.567 \]
Drunkard's Walk

Averaged Expectations (AE)

\[ \delta_{0.9}(M,N_0) \approx 0.64 \]
Drunkard's Walk

Averaged Expectations (AE)

\[ \delta_{0.9}(M,N_0) \approx 0.64 \]

\[ \delta_{0.9}(M,N_1) \approx 0.56 \]
Drunkard's Walk

Averaged Expectations (AE)

\[ \delta_{0.9}(M,N_0) \approx 0.64 \]

\[ \delta_{0.9}(M,N_1) \approx 0.56 \]

\[ \delta_{0.9}(M,N_2) \approx 0.543 \]
Drunkard's Walk

Averaged Expectations (AE)

\[
\delta_{0.9}(M,N_0) \approx 0.64
\]

\[
\delta_{0.9}(M,N_1) \approx 0.56
\]

\[
\delta_{0.9}(M,N_2) \approx 0.543
\]

\[
\delta_{0.9}(M,N_3) \approx 0.540
\]
| Case          | $|M|$ | $k$ | $\lambda = 1$ | $\lambda = 0.8$ |
|--------------|-----|-----|----------------|-----------------|
|              |     |     | $\delta_{\lambda}$-init | $\delta_{\lambda}$-final | # | time | $\delta_{\lambda}$-init | $\delta_{\lambda}$-final | # | time |
| IPv4 (AM)    | 23  | 5   | 0.775          | 0.054           | 3 | 4.8  | 0.576          | 0.025           | 3 | 4.8  |
|              | 53  | 5   | 0.856          | 0.062           | 3 | 25.7 | 0.667          | 0.029           | 3 | 25.9 |
|              | 103 | 5   | 0.923          | 0.067           | 3 | 116.3| 0.734          | 0.035           | 3 | 116.3|
|              | 203 | 6   | –              | –               | – | TO   | –              | –               | – | TO   |
| IPv4 (AE)    | 23  | 5   | 0.775          | 0.109           | 2 | 2.7  | 0.576          | 0.049           | 3 | 4.2  |
|              | 53  | 5   | 0.856          | 0.110           | 2 | 14.2 | 0.667          | 0.049           | 3 | 21.8 |
|              | 103 | 5   | 0.923          | 0.110           | 2 | 67.1 | 0.734          | 0.049           | 3 | 100.4|
|              | 203 | 6   | –              | –               | – | TO   | –              | –               | – | TO   |
| DrkW (AM)    | 39  | 7   | 0.565          | 0.466           | 14| 259.3| 0.432          | 0.323           | 14| 252.8|
|              | 49  | 7   | 0.568          | 0.460           | 14| 453.7| 0.433          | 0.322           | 14| 420.5|
|              | 59  | 8   | 0.646          | –               | – | TO   | 0.423          | –               | – | TO   |
| DrkW (AE)    | 39  | 7   | 0.565          | 0.435           | 11| 156.6| 0.432          | 0.321           | 2 | 28.6 |
|              | 49  | 7   | 0.568          | 0.434           | 10| 247.7| 0.433          | 0.316           | 2 | 46.2 |
|              | 59  | 8   | 0.646          | 0.435           | 10| 588.9| 0.423          | 0.309           | 2 | 115.7|

**Table 1.** Comparison of the performance of EM algorithm on the IPv4 zeroconf protocol and the classic Drunkard's Walk w.r.t. the heuristics AM and AE.
What we have seen

Theoretical Results
Metric-based state space reduction for MCs
1. Closest Bounded Approximant (CBA) encoded as a bilinear program
2. Bounded Approximant (BA) PSPACE & NP-hard for all $\lambda \in (0,1]$ 
3. Significant Bounded Approximant (SBA) NP-complete for $\lambda = 1$

Practical Results
We proposed an EM-like method to obtain a sub-optimal approximants
Previous work

• On-the-Fly Exact Computation of Bisimilarity Distances - TACAS 2013

• The BisimDIST Library: Efficient Computation of Bisimilarity Distances for Markovian Models - QEST 2013

• Computing Behavioral Distances, Compositionally - MFCS 2013

• Converging from Branching to Linear Metrics on Markov Chains - ICTAC 2015

• On the Metric-based Approximate Minimization of Markov Chains - ICALP 2017
Future Work

- Is BA-$\lambda$ SUM-OF-SQUARE-ROOTS-hard?
- Can we obtain a real/better EM-heuristics?
- What about different models/distances?